By the Numbers

Volume 15, Number 2 The Newsletter of the SABR Statistical Analysis Committee May, 2005

Summary

Academic Research: Ignoring Data to Improve Statistical Prediction

Charlie Pavitt

This academic study finds that in one certain analysis of baseball data, removing players with few plate appearances increased predictions of performance. Here, the author reviews that study, and suggests implications for sabermetric research.

This is one of a series of reviews of sabermetric articles published in academic journals. It is part of a project of mine to collect and catalog sabermetric research, and I would appreciate learning of and receiving copies of any studies of which I am unaware. Please visit the Statistical Baseball Research Bibliography at its new location www.udel.edu/communication/pavitt/biblioexplan.htm . Use it for your research, and let me know what is missing.

Thomas A. Timmerman, Missing Persons in the Study of Groups, Journal of Organizational Behavior, 2005, Volume 26, pp 21-36

This is a strange study, in that its implications for baseball researchers are much different than those for the author. Timmerman is no stranger to baseball research; this is the third baseball-oriented study of his of which I am aware, and the second that I have reviewed for BTN.

Timmerman's concern is methodological -- the impact of missing data on the measurement of group-level phenomena. For example, when some members of organizational groups fail to fill out surveys about their work experience, to what extent are statistical conclusions about their group more prone to both random and systematic error? Timmerman used basketball and baseball data for the same reason as other organizational researchers; it is trustworthy and easy to obtain.

I will stick with the baseball analyses here. Using 100 years of data (1903 through 2002), Timmerman regressed team winning percentage on team OPS, team ERA, and the team average of its players' years of major league experience; the respective betas for these three predictors were .316, -.313, and .248.

Timmerman apparently believes that the correlation between a team's winning percentage and its players' MLB experience "is consistent with the idea that team experience leads to knowledge, skills, and abilities *about teamwork* that may lead to better team performance" [emphasis his]. I doubt it. It seems to me that this association is consistent with the facts that better teams have better players and that better players have longer careers. Second, Timmerman randomly removed different proportions of player data from his data set (10%, 20%, and so on), finding that the greater proportion removed, the lower the associations; with 90% of the players removed, the regression coefficients were only .128 for team OPS, -.097 for team ERA, and .131 for years of experience.

Third, and of particular interest to us, Timmerman systematically removed different proportions of player data based on the number of plate appearances and innings pitched in a given season; i.e., 10% least active, 20% least active, and so on, under the rationale that the inclusion of data from organizational

In this issue

members who have contributed little to group goals improves prediction of group performance. Indeed, prediction of team performance from player performance had an inverted-U relationship with proportion of players removed, at first improving markedly with removal of less active players from the data, leveling off with the 40% least active removal, then starting to decrease at 60%.

At their peak, with the 50% least active players and pitchers removed, the regression coefficients were .763 for team OPS and -.794 for team ERA. There was no analogous effect for experience.

I see the following implication of this finding for our research: When examining player performance, we often guesstimate how many plate appearances or innings pitched should serve as a cut-off for including a player in a data set; for example, one might consider 200 plate appearances as constituting the criterion for reasonably regular play. Although it would be unfair to say that the cut-offs we choose are arbitrary, they are not as principled as we might like. Perhaps the 50% rule implied by these findings might serve as a defensible criterion for sufficient play.

Charlie Pavitt, 812 Carter Road, Rockville, MD, 20852, chazzq@udel.edu ♦

Submissions

Phil Birnbaum, Editor

Submissions to *By the Numbers* are, of course, encouraged. Articles should be concise (though not necessarily short), and pertain to statistical analysis of baseball. Letters to the Editor, original research, opinions, summaries of existing research, criticism, and reviews of other work are all welcome.

Articles should be submitted in electronic form, either by e-mail or on CD. I can read most word processor formats. If you send charts, please send them in word processor form rather than in spreadsheet. Unless you specify otherwise, I may send your work to others for comment (i.e., informal peer review).

If your submission discusses a previous BTN article, the author of that article may be asked to reply briefly in the same issue in which your letter or article appears.

I usually edit for spelling and grammar. If you can (and I understand it isn't always possible), try to format your article roughly the same way BTN does.

I will acknowledge all articles upon receipt, and will try, within a reasonable time, to let you know if your submission is accepted.

Send submissions to:
Phil Birnbaum
88 Westpointe Cres., Nepean, ON, Canada, K2G 5Y8
birnbaum@sympatico.ca

Clutch Hitting: Recent Studies

Phil Birnbaum

Several articles on clutch hitting have appeared recently; they are reviewed here by the author.

The clutch-hitting debate continues; several articles on the topic have appeared within the last couple of months, at least two of them being thorough, large-scale investigations.

Bill James, "Mapping the Fog," and Phil Birnbaum, "Response to 'Mapping the Fog'"

In the previous issue of BTN, Jim Albert and I (separately) wrote articles commenting on Bill James' already-famous "Underestimating the Fog," which appeared in this year's *Baseball Research Journal*.

Following our comments, Bill wrote a rebuttal to our articles (but mostly to mine), which he distributed to the SABR Statistical Analysis Committee mailing list. Among his many arguments are that, under certain assumptions about clutch hitting – for example, if only 20% of players exhibit it -- the Cramer test (which my article defended) will fail to find it. Also, he claims that my methodology is not statistically legitimate.

My own article argues that Bill used a much weaker version of the Cramer test, and the actual Cramer test would *not* fail to find the result; that my methodology is legitimate; and that James' main argument – "when you get random data, all you can conclude is that your study has failed" – is not generally correct.

The rebuttals are available from the authors, or at the "Baseball Think Factory" web page below.

Both articles: http://www.baseballthinkfactory.org/files/main/article/discussing_the_fog

Elan Fuld, "Clutch and Choke Hitters in Major League Baseball: Romantic Myth or Empirical Fact"

This study, by a University of Pennsylvania student, is an analysis of clutch-hitting data derived from all play-by-play data recently available at Retrosheet. Unlike other clutch-hitting studies, which divide plate appearances into "clutch" and "non-clutch," Fuld ranks each at-bat on a sliding scale of "clutchness", based on its effect on a team's probability of winning. He then runs a regression of performance versus clutchness, and searches for statistical significance.

This study has achieved a fair bit of publicity in the mainstream press[1], perhaps prompted by a news release from the University of Pennsylvania itself. "Elan Fuld has proven it: clutch hitters really do exist," says the release, and that conclusion is echoed by most of the press articles reporting the study.

But a closer look at Fuld's results seems to suggest the opposite conclusion – that Fuld has proven that clutch hitters do *not* exist.

The main issue is sacrifice flies. Fuld wasn't sure how to classify those, so he tried three different ways – counting them as outs, counting them as equivalent to walks, or not counting them at all.

"When I included sacrifice flies [as walks] in the analysis," Fuld writes, "there was overwhelming evidence that there were clutch hitters." But counting them as outs, Fuld found *no* statistically-significant evidence of clutch hitting.

[1] You can Google "Elan Fuld Clutch" to get a large assortment of links.

And it seems obvious that counting them as outs is the correct method for this study. Sacrifice flies occur only when a runner is on third base, and those situations are much more likely to be clutch. Counting them as walks instead of outs artificially increases a player's statistics, and those artificial increases accrue disproportionately on the clutch side. The same fly ball that is counted as a negative when non-clutch is counted as positive in the clutch, which gives all players the appearance of clutch hitting. That is, even if all players perform *exactly the same* in the clutch as in normal situations, the performance is counted differently in favor of clutch.

For a back-of-the-envelope calculation, suppose an average player hits 4 SF a year, 3 of which come among 250 clutch plate appearances, the last one coming among 250 non-clutch PA. Counting all four of them as walks means that the player's on-base percentage winds up .008 higher in the clutch, for no other reason than the artificial scoring method.

A difference of 8 points may not seem like much, but a study based on this much data is going to be very senstive to small differences. This can be seen on page 43 of Fuld's paper, where, without exception, he finds significance only in cases where sacrifice flies are counted as walks.

Two more of Fuld's findings support this conclusion. First, he finds many more clutch hitters than choke hitters, which you would expect if you give clutch hitting stats an artificial boost. Second, when he lists several tables of players who were significant clutch or choke hitters, he finds many such hitters when counting SFs as walks, but *none* when counting them as outs.

I would argue that Fuld has performed an exceptionally detailed and thorough study, but that the results just don't support his conclusion.

Which, in a way, is fortunate. If Fuld concluded that clutch hitting does not exist, there probably wouldn't have been any publicity – and we wouldn't have found out about his excellent new study.

Study: http://www.soapboxincyberspace.com/clutch%20writeup1.%2071.pdf
Press Release: http://www.upenn.edu/pennnews/article.php?id=793
For other Fuld links, see Cyril Morong's page: http://www.geocities.com/cyrilmorong@sbcglobal.net/ClutchLinks2.htm

Tom Ruane, "In Search of Clutch Hitting"

In the March, 1990 issue of "By the Numbers," Pete Palmer searched for the existence of clutch hitting using this logic: if clutch hitting didn't exist, players' clutch hitting stats would follow the same normal distribution as if you chose clutch-hitting numbers randomly. But if clutch hitting *did* exist, the distribution would be spread out much farther than random, as the good and bad clutch hitters would add datapoints to the far extremes of the bell curve.

Using 10 years' worth of data from the *Elias Baseball Analyst*, Palmer concluded that the extremes of the bell-curve were "just as expected" from random data, and that clutch hitting was not shown to exist.

Fifteen years later, Tom Ruane ran a similar study, based on all of Retrosheet's available data (and also, I believe, data from the 1990s not publicly available from the site). His study is exquisitely detailed, anticipating almost every nitpick that could be raised by doubters.

He adjusts for the fact that players hit better in different base-out situations. He counts all sacrifice flies as outs. He ignores walks, as they may be partially a function of defensive strategy. He considers the effects of handedness and platooning. He checks the quality of opposition pitching. And he looks at park factors.

His conclusion is the same as Palmer's – no evidence of clutch hitting. Ruane's curve of actual clutch hitting looks just like his simulated curve of random clutch hitting.

And, the Palmer/Ruane test should be powerful enough to find an effect localized among a small number of players. Ruane's study found that for their careers, there were 28 real-life players who hit more than 75 points (on-base plus batting average) better in the clutch; the simulation produced 27. If even, say, 20 players out of the 700 studied were signficantly clutch, the results would likely be far more outsized than 28-27.

This level of sensitivity wouldn't be as high for the Cramer test, which (as Bill James points out in "Mapping the Fog," reviewed above) would probably fail to find an effect localized among a very small number of players.

Ruane is careful to emphasize that his study isn't actual proof that clutch hitting doesn't exist, but argues that "the forces at work here, if they exist, must be awfully weak to so closely mimic random noise, and if they are really that inconsequential perhaps we could assume they don't exist without much loss of accuracy."

For its level of detail and its probable sensitivity, this study is probably the most authoritative of any clutch-hitting study done to date.

Study: http://www.retrosheet.org/Research/RuaneT/clutch_art.htm

----

Phil Birnbaum, birnbaum@sympatico.ca ♦

Informal Peer Review

The following committee members have volunteered to be contacted by other members for informal peer review of articles.

Please contact any of our volunteers on an as-needed basis - that is, if you want someone to look over your manuscript in advance, these people are willing. Of course, I'll be doing a bit of that too, but, as much as I'd like to, I don't have time to contact every contributor with detailed comments on their work. (I will get back to you on more serious issues, like if I don't understand part of your method or results.)

If you'd like to be added to the list, send your name, e-mail address, and areas of expertise (don't worry if you don't have any - I certainly don't), and you'll see your name in print next issue.

Expertise in "Statistics" below means "real" statistics, as opposed to baseball statistics - confidence intervals, testing, sampling, and so on.

Member	E-mail	Expertise
Ben Baumer	bbaumer@nymets.com	Statistics
Jim Box	jim.box@duke.edu	Statistics
Keith Carlson	kcsqrd@charter.net	General
Dan Evans	devans@seattlemariners.com	General
Rob Fabrizzio	rfabrizzio@bigfoot.com	Statistics
Larry Grasso	l.grasso@juno.com	Statistics
Tom Hanrahan	Han60Man@aol.com	Statistics
John Heer	jheer@walterhav.com	Proofreading
Dan Heisman	danheisman@comcast.net	General
Bill Johnson	firebee02@hotmail.com	Statistics
David Kaplan	dkaplan@UDel.Edu	Statistics (regression)
Keith Karcher	karcherk@earthlink.net	Statistics
Chris Leach	chrisleach@yahoo.com	General
Chris Long	clong@padres.com	Statistics
John Matthew IV	john.matthew@rogers.com	Apostrophes
Nicholas Miceli	nsmiceli@yahoo.com	Statistics
Duke Rankin	RankinD@montevallo.edu	Statistics
John Stryker	johns@mcfeely.interaccess.com	General
Tom Thress	TomThress@aol.com	Statistics (regression)
Joel Tscherne	Joel@tscherne.org	General
Dick Unruh	runruhjr@dtgnet.com	Proofreading
Steve Wang	scwang@fas.harvard.edu	Statistics

Correction to "Does Good Hitting Beat Good Pitching?"

Tom Hanrahan

In an article in a previous BTN, the author used an established relationship between slugging percentage and ERA to estimate SLG for pitchers with the lowest ERAs in the league. However, this introduced a selection bias, as deliberately selecting pitchers with the lowest ERAs is more likely to include pitchers who, for luck or other reasons, had SLGs higher than would normally be expected by ERA. Here, the author explains why this happens, and issues a caveat for his original study.

In "*Does Good Hitting Beat Good Pitching*?" (BTN, August 2001, and reprinted in "The Best of *By the Numbers*"), I measured actual batting results when "good" or "bad" hitters faced "good" or "bad" pitchers, and compared this against expected results using a standard mathematical relationship. I used batting averages for the comparison, and found that good pitching did not "beat" good hitting any more than one would expect, given the input data. I further attempted to compare on-base and slugging percentages.

However, I didn't have actual data for slugging percentage, so I estimated it. As I wrote in the original article,

> Data for pitchers' SLG allowed was not available. However, this can be derived from the opponents' OBA and ERA. The league averages for the period 1984-1996, weighted by player-years active in study, were OBA = .3274, SLG = .3988, and ERA = 4.074. From this we can relate ERA to the offensive components by the formula ERA = OBA * SLG * 31.20 for this period, using the accepted structure for Runs Created. Since this formula works for leagues (the RMS error in computing ERA for individual seasons using the formula was 0.08, or 2% of the ERA), I have assumed it works for individual pitchers.

Using this method, the article calculated that good hitters should have slugged .4287 against good hitters. But they actually did slug .4545. I therefore concluded that since actual slugging was 26 points higher than expected (.4545 versus .4287), "good" sluggers might be able to take advantage of "good" pitchers in some unforeseen fashion.

But.

I now see that there is a flaw in the way I derived the pitchers' "slugging allowed," which quite likely drove the original study to an erroneous conclusion.

When I attempted to measure on-base and slugging averages, I had to estimate pitchers' slugging allowed through use of their ERAs, since the data were not directly available. While the relationship between OBA, SLG and ERA is well defined, I inadvertently may have introduced a 'selection bias'. By defining "good" pitchers by ERA (lower than 3.53), I could have chosen a subset of pitchers with lower-than-expected-ERAs than their actual "slugging allowed" would indicate – obtained either through inducing extra GIDPs, stranding runners, picking base runners off, or merely luck. Hence, in my group of "good" pitchers, attempting to measure their SLG allowed by using their ERA in a formula could have been wrong.

This type of error occurs frequently (unfortunately). I saw it years ago when a study (which I will paraphrase from memory) concluded that it took an average of 8 runs to make an extra win. The study found all teams with records 10 games above .500 (91-71), found that these teams averaged scoring 80 more runs than they allowed, divided 80 by 10 and determined 8 runs = 1 win. The problem is, by using W-L records to parse the data, many more "lucky" teams with poor RS/RA ratios that won 91 games were included than those "unlucky" teams with superior RS/RA ratios that only won 91 games, because there simply weren't as many superior teams. When the study was restructured to look for all teams that outscored their opponents by 80 runs and *then* find the W-L records, it was determined that these teams finished an average of 89-73, and so 80 runs only equals 8 extra wins; 10 runs = 1 win, which is much closer to the truth in most cases.

So, in your own research, be aware of this, or we shall all accuse you of pulling a Hanrahan in the future.

Tom Hanrahan, Han60Man@aol.com. ♦

Get Your Own Copy

If you're not a member of the Statistical Analysis Committee, you're probably reading a friend's copy of this issue of BTN, or perhaps you paid for a copy through the SABR office.

If that's the case, you might want to consider joining the Committee, which will get you an automatic subscription to BTN. There are no extra charges (besides the regular SABR membership fee) or obligations – just an interest in the statistical analysis of baseball.

The easiest way to join the committee is to visit http://members.sabr.org, click on "my SABR," then "committees and regionals," then "add new" committee. Add the Statistical Analysis Committee, and you're done. You will be informed when new issues are available for downloading from the internet.

If you would like more information, send an e-mail (preferably with your snail mail address for our records) to Neal Traven, at beisbol@alumni.pitt.edu. If you don't have internet access, we will send you BTN by mail; write to Neal at 4317 Dayton Ave. N. #201, Seattle, WA, 98103-7154.

Is OBP Really Worth Three Times as Much as SLG?

Phil Birnbaum

Michael Lewis's book has Paul Depodesta arguing that a point of OBP is worth three times as much as SLG. Is that correct? Here, the author devises a way to measure the ratio between the two.

Michael Lewis's "Moneyball" has Paul DePodesta arguing that, on average, a point of On-Base Percentage is worth three times as much as a point of Slugging Percentage. Is that true?

To answer that question, let's start with a typical, average-type batting line – the entire 1987 American League.[1]

	AB	H	2B	3B	HR	BB	OBP	SLG
1987 AL	77819	20620	3667	461	2634	7812	.33203	.42549

What I'm going to do is this: First, I'll create a batting line that's exactly the same as this one, except that the SLG is one point higher. Then, I'm going to create a clone where only the OBP is one point higher. Once we've done that, we can compare the two increases, and see how much more one is worth than the other.

SLG

As I said, we want to leave OBP the same, but bump up slugging percentage by exactly one point (.001). The only way to do that is to leave the number of hits and walks the same, but bump up total bases – that is, extra bases – by exactly 77.819.

One way to do that is to change 77.819 singles into doubles.

	AB	H	2B	3B	HR	BB	OBP	SLG
Original	77819	20620	3667	461	2634	7812	.33203	.42549
New	77819	20620	3744.819	461	2634	7812	.33203	.42649

By doing that, we've converted 77.819 singles, each with linear weight .46, to 77.819 doubles, each with linear weight .80. Each of those conversions, then, is worth the difference between .46 and .80 – that's +.34 runs each.

Had we converted doubles (.80) to triples (1.02), they would be worth +.22 runs each. And if we converted triples (1.02) to home runs (1.40), they would be worth +.38 runs each. Taking a rough average of these three values (+.34, +.22, +.38), and keeping in mind that triples are rare, we can say, roughly, that each additional slugging base is worth about +.35 runs.

Multiplying that .35 by the 77.819 extra bases gives us a total of about 27 runs.

- If the 1987 AL had a slugging percentage one point higher, it would have scored about 27 additional runs.

OBP – Walks

Now, suppose we want to leave SLG the same, but increase OBP by one point. There are two ways to do that – through a hit, or through a walk. We'll start with the walk.

There were 85,631 plate appearances (ignoring HBP for now) in the 1987 American League. So, converting 85.631 outs to walks will increase OBP by exactly 1 point:

[1] You'll notice I'm ignoring sacrifice flies and HBP in the calculation of OBP, but this simplification won't matter much in the final result. Neither will the arbitrary choice of the 1987 AL as our starting point.

	AB	H	2B	3B	HR	BB	OBP	SLG
Original	77819	20620	3667	461	2634	7812	.33203	.42549
Step 1	77733.369	20620	3667	461	2634	7897.631	.33303	.42596

But converting the outs to walks reduced the number of at-bats – and, in so doing, raised the SLG from .42549 to .42596. To bring the SLG back down, we need to lose 36.435 total bases. We can do this by converting 36.435 doubles to singles:

	AB	H	2B	3B	HR	BB	OBP	SLG
Original	77819	20620	3667	461	2634	7812	.33203	.42549
Step 1	77733.369	20620	3667	461	2634	7897.631	.33303	.42596
New	77733.369	20620	3630.565	461	2634	7897.631	.33303	.42549

(Again, we could have converted triples to doubles, or Home Runs to triples – but we'll illustrate the difference with doubles, and again use the blended linear weight value of +.35.)

We now have exactly what we were looking for – a one point increase of OBP, with no change to SLG. How much is that worth?

The 36.435 bases lost at .35 runs each is a loss of 12.75 runs. The extra 85.631 walks, at a linear weight of +.33 runs, is a gain of 28.26 runs. The 85.631 outs no longer made, at a linear weight of .25 runs each, is a gain of 21.41 runs.

36.435 extra bases lost	.35 runs each	-12.75 runs
85.631 walks gained	.33 runs each	+28.26 runs
85.631 outs saved	.25 runs each	+21.41 runs
TOTAL		+36.92 runs

The total, then, is a gain of 36.92 runs.

- If the 1987 AL gained a point of OBP through walks, it would have scored about 37 additional runs.

OBP – Hits

The other way to increase OBP is through base hits. And so, we can convert 85.631 outs to singles instead of walks:

	AB	H	2B	3B	HR	BB	OBP	SLG
Original	77819	20620	3667	461	2634	7812	.33203	.42549
Step 1	77819	20705.631	3667	461	2634	7812	.33303	.42659

But the extra singles also increase SLG. To bring SLG back down to the original value, we'll need to subtract 85.631 extra bases, which we'll do by converting doubles to singles:

	AB	H	2B	3B	HR	BB	OBP	SLG
Original	77819	20620	3667	461	2634	7812	.33203	.42549
Step 1	77819	20705.631	3667	461	2634	7812	.33303	.42659
Final	77819	20705.631	3581.369	461	2634	7812	.33303	.42549

The line "Final" is now what we want – same SLG, and one point higher OBP.

Repeating the calculation for hits gives:

85.631 extra bases lost	.35 runs each	-29.97 runs
85.631 singles gained	.46 runs each	+39.39 runs
85.631 outs saved	.25 runs each	+21.41 runs
TOTAL		+30.83 runs

- If the 1987 AL gained a point of OBP through hits, it would have scored about 31 additional runs.

OBP – Overall

So a point of OBP was worth 37 runs when achieved through walks, but only 31 runs when achieved through hits. Weighting the average towards the hit value (since hits are about 2.6 times as common as walks), we can say that a point of OBP was worth about 32.6 runs. Taking into account HBP, which we didn't count but are equivalent to walks, we'll raise that to 33 runs.

- If the 1987 AL gained a point of OBP through a typical combination of walks and hits, it would have scored about 33 additional runs.

The Relative Weight

So a point of SLG was worth about 27 runs, and a point of OBP was worth 33 runs. Dividing 33 by 27 gives us approximately 1.2, which means

- A point of OBP is worth 1.2 times as much as a point of SLG.

Applications

Suppose we have two players, both of which have an OPS of .800. Player A has an OBP of .350 and an SLG of .450, while player B has an OBP of .360 and an SLG of .440. We can conclude that player B is slightly better than player A.

How much better? We don't know offhand. We do know that player B's 10 points of OBP are worth about 20% more than player A's 10 points of SLG. But how much is 20% more? One run? Two runs? Ten runs? There's no easy formula to convert OPS to runs, so how can we convert "20% of a 10 point difference in SLG" to runs?

We can't, and we wouldn't want to. OPS is only a rough indicator of offensive value; trying to use OPS for this kind of valuation is like trying to sharpen a plastic butter knife. There are many other, better statistics that we can use to compare the two players – Runs Created, Linear Weights, and Extrapolated Runs are fully suited to this kind of task.

Notes

- Using some other batting line other than the 1987 American League would change the results slightly, but not significantly. The same for using different linear weights; the weights we used were for lower-scoring offenses than the 1987 AL, but the results would not be significantly impacted if we used slightly higher weights.

- We have ignored the fact that SFs negatively impact OBP; again, this should not have significantly affected the results.

- In trying to bump up OBP with hits, we added only singles. We could, instead, have added extra-base hits. But the result would have been the same. If we had added doubles, for instance, we would have had to bump those same doubles back down to singles to reset SLG, giving the same result.

Phil Birnbaum, birnbaum@sympatico.ca ♦